REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
23-02-2010	Final Report	15-Aug-2006 - 14-Aug-2009

4. TITLE AND SUBTITLE
Network-Centric Distributed Signal Processing

5a. CONTRACT NUMBER
W911NF-06-1-0346

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

6. AUTHORS
Lang Tong

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
Cornell University
Office of Sponsored Programs
Cornell University
Ithaca, NY 14853 -2801

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
50599-NS.1

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
This project investigates distributed signal processing in the context of sensor enhanced mobile ad hoc networks. The objective of the project is twofold. First, the research aims to create a new framework for Network Centric Signal Processing that facilitates a joint design of communications,networking, and signal processing optimized for specific applications. Second, for sensor networks deployed for detection and estimation, this research develops new multiaccess communication strategies, networking protocols (medium access control and routing), and

15. SUBJECT TERMS
Sensor networks, distributed statistical inference, wireless communications

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
UU	UU	UU	UU		Lang Tong
					19b. TELEPHONE NUMBER
					607-255-3900

Standard Form 298 (Rev 8/98)
Prescribed by ANSI Std. Z39.18

Network-Centric Distributed Signal Processing

Final Report Submitted to

The US Army Research Office

Contract No. ARO-W911NF-06-1-0346

Sept 2006–Sept 2009

Submitted to the Army Research Office

Lang Tong, PI

School of Electrical and Computer Engineering

Cornell University, Ithaca, NY 14853

Email: `ltong@ece.cornell.edu`

URL: `people.ece.cornell.edu/ltong`

Contents

1 Overview

This project investigates distributed signal processing in the context of sensor enhanced mobile ad hoc networks. The objective of the project is twofold. First, the research aims to create a new framework for Network Centric Signal Processing that facilitates a joint design of communications,networking, and signal processing optimized for specific applications. Second, for sensor networks deployed for detection and estimation, this research develops new multiaccess communication strategies, networking protocols (medium access control and routing), and centralized and decentralized signal detection and estimation algorithms.

The approach adopted in this project is based on a cross-layer design paradigm that integrates application layer performance measure with physical layer signaling, medium access protocols, and network-wide route discovery and route selections. Coupled with numerical simulations and experimental testing, the project applies analytical techniques drawn from the theory of large deviations, asymptotic statistics, graphical models, and combinatorial optimization.

This project has resulted in 8 journal publications, 6 conference publications, and 3 book/book chapters. One of the journal papers [1] supported by this grant received the *2008 IEEE Signal Processing Society Young Author Best Paper Award*.

This project investigates four related subtopics on network centric signal processing as highlighted below.

- *Energy scaling laws for distributed inference in random networks [2, 3]:* We investigate the fundamental problem of how energy consumption of a random sensor network scales with the size of the network. Specifically, we are interested in establishing the energy scaling laws that characterizes the growth of energy consumption per node when optimal decisions are to be made at the fusion center. We showed that, for hypothesis testing problems, conventional internet-style routing will not lead to scalable fusion, and in-network computation is essential for any scalable fusion.

- *Optimal sensor deployment and activation [4, 5]:* Energy consumption is a major limiting factor for sensor network design. For large sensor networks, we consider the problem of optimal random activation and placement of sensors. In general, the problem of energy consumption must be addressed simultaneously with network performance. To this end, we consider the probability of detection as a measure of performance and investigate the tradeoff between accurate inference and energy consumption in distributed detection.

- *Distributed inference over networks: [6, 7, 1, 8, 9, 10, 11]* We consider the problem of data fusion in a wireless network. In particular, we investigate the impact of network constraints such as bandwidth allocation and channel impairments such as fading and interference on fusion performance.

- *Coded cooperative sensor fusion [12, 13, 14]:* We investigate coding strategies for distributed inference when sensors are either misinformed for the message or are attacked by adversaries.

For robust sensor fusion, it is therefore necessary to consider coding techniques.

2 Summary of Major Findings

2.1 Energy Scaling Laws for Distributed Inference in Random Networks

We consider the problem of distributed statistical inference via a network of randomly located sensors, each taking measurements and transporting the locally processed data to a fusion center. The fusion center then makes an inference about the underlying phenomenon based on data collected from individual sensors.

For statistical inference using wireless sensor networks, energy consumption is one of the most important design factors. The transmission power required for a receiver distance d away to have a certain signal-to-noise ratio (SNR) scales with the distance to the power ν where $2 \leq \nu \leq 6$ is the path loss. Therefore, the cost of moving data from sensor locations to the fusion center either through direct transmissions or multihop forwarding significantly affects the lifetime of the network.

In this work, we allow data aggregation at intermediate nodes, but require that the fusion center achieves the same inference performance as if all raw observations were collected. We assume that the underlying hypotheses can be modeled as Markov random fields and investigate energy scaling laws.

Given sensor locations and possibly correlated sensor measurements, finding the minimum energy fusion policy is in general NP-hard and hence, intractable. We will establish upper and lower bounds on the fusion energy of the optimal scheme and analyze its scaling behavior. The lower bound is achieved by a minimum spanning tree fusion scheme, which is shown to be optimal when observations are statistically independent under both hypotheses. The upper bound is established through a specific suboptimal fusion scheme, referred to as Data Fusion over Markov Random Field (DFMRF). DFMRF becomes optimal for conditionally independent observations, and for certain spatial dependencies between sensor measurements of practical significance (e.g., nearest neighbor graph); it has an approximation ratio 2, i.e., it costs no more than twice the cost of the optimal fusion scheme, independent of the size of the network.

We then establish a number of asymptotic properties of DFMRF, including the scalability of DFMRF, its performance bounds, and the approximation ratio with respect to the optimal fusion policy when the sensor measurements have dependencies described by a k-nearest neighbor graph or a disk graph (continuum percolation).

We also provide a precise characterization of the scaling bounds as a function of sensor node density and sensor distribution. These asymptotic bounds for DFMRF, in turn, are also applicable to the optimal fusion scheme. Hence, we use the DFMRF scheme as a vehicle to establish scaling laws for optimal fusion. Additionally, we use the expressions for scaling bounds to optimize the distribution of the sensor placements. For conditionally independent measurements and for correlated measurements with k-nearest neighbor dependency graph, we show that the uniform distribution minimizes the scaling bounds over all i.i.d placements.

To the best of our knowledge, our results are the first to establish the scalability of data fusion for certain correlation structures1 of the sensor measurements. The use of an energy scaling law for the design of sensor placement is new and has direct engineering implications. The heuristic policy DFMRF first appeared [?] and is made precise here with detailed asymptotic analysis using the weak law of large numbers for stabilizing graph functionals.

2.2 Optimal Sensor Deployment

We investigate optimal strategies for sensor deployments in the context of distributed sensing and inference. In particular, we are interested in the optimal selection of sensor density and sensor distributions, which has direct implications of sensor network architectures, sensor activation/sleeping strategies.

The deployment of sensors is the first step in establishing a network. It influences the performance of the network, including the energy consumed in routing data. However, designing optimal deployment strategies requires finding optimal locations for all nodes and is not feasible for a large network. In contrast, deploying nodes randomly but according to an optimal density may be more tractable. In this work, we assume that the nodes are placed IID according to certain distribution with an optimized sensor density.

In a network of sensors measuring a correlated signal field, the node density influences the extent of correlation among the measurements and thereby detection performance. Moreover, the energy required to route data typically depends on the inter-node distance, and in turn, the node density. Hence, we can design an optimal density that achieves the best tradeoff between the energy consumed in routing data and the resulting detection performance. Of course, the optimal density depends on the routing protocols employed.

In the classical approach, a layered architecture separates the design of routing from the application. For energy constrained networks, application-specific routing may offer better tradeoffs, and we employ one such routing scheme. To characterize the detection performance, we consider the Neyman-Pearson (NP) error exponent. Our objective is to find an optimal node density that maximizes the detection error exponent D, under a constraint on per-node energy consumption.

Note that both the energy constraint and detection performance are asymptotic in the number of nodes. We address the following questions: does an optimal density exist? And if so, what is its value? Is it one of the extremes, viz., zero or infinity? This is an important question, since if the optimal node density is either zero or infinity, then we can simply place the nodes in as small/large an observation area as possible. On the other hand, if this is not the case, then we need to deploy the nodes, based on the optimal density.

We assume the presence of correlation under the alternative hypothesis H1 through the Gauss-Markov random field (GMRF) model. Again, assuming an uniform signal field model (same measurement variance at every node) and a correlation function decaying with distance, we study the effect of node density on the error exponent. To this end, we unify the results from our previous works which independently characterize the detection performance and routing for a GMRF.

Intuitively, when both the hypotheses have the same measurement variance, they can be distinguished only by the presence of correlation under H1. Correlation is maximized when all the nodes are clustered close to one another, since correlation decays with distance. Hence, ignoring the energy constraint, the optimal density should be infinite. We prove that this is indeed true.

For the general case, when the measurement variances are different, the behavior is decided by the ratio of the variances under H1 and H0. We show that the optimal density is infinite when the variance ratio is below a threshold value. Moreover, imposing any feasible routing-energy constraint does not change the optimal density. We provide the expression for the threshold on the variance ratio. On the other hand, above this threshold the optimal density is characterized by the energy constraint.

2.3 Statistical Inference over networks

Classical distributed inference focuses on the design of local quantization and fusion rules. The design of large wireless-sensor networks (WSN) must deal with challenges beyond the optimization of such decision rules, and the optimization must take into account communication and networking aspects of the problem: the presence of fading, the multiuser interference, and the sharing of limitted bandwidth.

We investigate the problem of distributed statistical inference over a large wireless sensor network. Our goal is to develop a network-centric approach by focusing on multiaccess aspect of the problem. The communication scheme we employ is the so-called Type-Based Random Access (TBRA) in which sensors transmit probabilistically using a set of orthogonal waveforms keyed to their measurement. Specifically, sensors with the same data value transmit using the same waveform on a multiaccess fading channel. The use of orthogonal waveforms eliminates interference among users with different data values and makes it possible to have coherent combining of transmissions in the absence of fading. The behavior is more complicated in presence of fading, however, since there is a possibility of cancellation among the signals.

For the problem of distributed estimation, we have analyzed the Bayesian estimation performance in terms of the expected Fisher information normalized by the transmission rate of the sensors. Under a constraint on the expected transmission energy, an optimal spatio-temporal allocation scheme that maximizes the performance metric is characterized. It is shown that the metric is crucially dependent on the fading parameter known as the channel coherence index. For channels with low coherence indices, sensor transmissions tend to cancel each other, and there exists an optimal mean transmission rate that maximizes the performance metric. On the other hand, for channels with high coherence indices, there should be as many simultaneous transmissions as allowed by the network. The presence of a critical coherence index, where the change from one behavior to another occurs, is established.

For distributed detection, given the fixed budget transmission energy, we obtain a characterization inference performance—error exponent for detection, mean square error (MSE) for estimation. This characterization leads to an optimal energy space-time distribution. We show that for chan-

nel with high coherent index, simultaneous transmission is optimal whereas for channels with low coherent index, there exists an optimal mean transmission rate in each time slot. This mean transmission rate translates to an optimal wake up strategy. Although optimal mean transmission rate is in general difficult to evaluate, we are able to derive an approximate solution based on a special version of central limit theorem.

For many applications, a sensor network operates in three phases. In the sensing phase, sensors take measurements that form a snap-shot of the signal field at a particular time. The measurements are stored locally. The second phase is information retrieval in which data are collected from individual sensors. The last phase is information processing in which data from sensors are processed centrally with a specific performance metric.

An appropriate network architecture for such applications is Sensor Networks with Mobile Access (SENMA)which has two types of nodes: low-power low-complexity sensors randomly deployed in large number and a few powerful mobile access points that communicate with sensors. Sensors communicate to the mobile access points (APs) directly. The use of mobile APs enable data collections from specific areas of the network, either by scheduling or by random access.

We have focused on the latter two phases in SENMA: information retrieval and processing. Specifically, we examine the effect of medium access control (MAC) for information retrieval on information processing in a network in which sensors are unreliable, possibly with low duty cycle, and subject to outage. In such cases, it is not possible for mobile access points to collect data from all sensors. We consider two types of MAC: deterministic scheduling that schedules transmissions from specific sensor locations, and random access that collects packets randomly from the sensor field.

For information processing, packets collected by mobile APs form a sampled signal field either randomly or deterministically with a specific pattern. Two factors affect the performance of information processing: the sampling pattern and the MAC throughput. The former tells us how the signal field is sampled, and the latter provides the amount of samples we can obtain during a collection time using a specific MAC. While it may appear obvious that collecting data from optimally chosen locations using (centralized) deterministic scheduling gives better performance, there are several nontrivial practical complications. For networks with finite sensor density, the scheduling scheme must have a finite scheduling resolution. By this we mean that, because the probability that a randomly deployed sensor exists at a particular location is zero, the deterministic scheduling protocol must be modified to schedule, not a sensor at a particular location, but sensor(s) in the neighborhood of a location. Even with such a modification, there is still a none-zero probability that the scheduled transmission location is void of sensors, or the batteries of those scheduled sensors have run out at the time of data collection.

The possibility of sensor outage brings the question of whether a deterministic MAC is sufficiently robust to practical imperfections. In this work, we compare deterministic scheduling and random access for the application of reconstructing a signal field with Poisson distributed sensors of finite density. To make the problem more tractable, we study the performance in a one dimensional signal field, which provides insight into the two-dimensional problem. A minimum mean square

error (MMSE) estimator is used for signal reconstruction, and the performance metric is the expected maximum MSE of the entire signal field. A key parameter is the sensor outage probability which is a function of sensor density, sensor duty cycle, and scheduling resolution*which gives the probability that there is no active sensor within the scheduled resolution interval. We show that, for large networks, there is an outage probability threshold beyond which the deterministic central scheduling is inferior to distributed random access.

2.4 Coded Cooperative Sensor Fusion

We investigate here the problem of extracting information from a large sensor network in which sensors cooperatively deliver messages to a mobile access point using a common codebook. If all collaborating sensors have agreed on a message, each sensor may transmit some part of the codeword that corresponds to the agreed message according to some schedule. In such a way, errors caused by channel noise can be corrected at the access point. Between the access point and the cooperative sensor network, there is a maximum achievable rate C_0 of information retrieval, below which the detection error at the access point can be made arbitrarily small by making the codeword length sufficiently long. But for large sensor networks in which sensors are distributed geographically and inexpensive with limited transmission and processing power, making all sensors agree on a common message is not easy. It is thus inevitable that some sensors will be mistaken on the message that is to be delivered cooperatively. Not knowing their mistakes, these misinformed sensors will transmit signals corresponding to the wrong codewords. The capacity of the sensor network with misinformed nodes is the maximum achievable rate C of information retrieval in the presence of not only channel noise but also sensor mistakes. Referred to as the capacity of the network with misinformed sensors, C is expected to be less than C_0.

In this work, we investigate practical coding schemes information retrieval assuming that the codebook used has a rate R below the capacity. For the fixed code rate R, we are interested in designing the parameter k of the stay-k scheduling so that the decoder has the fastest decay rate of error probability. To this end, we first derive the random coding error exponent as a function of rate R and the scheduling parameter k. We show next that, for any $R < C$, the error exponent approaches to zero as k 8, which means that, in contrast to the capacity achieving strategy, there is an optimal k* that the access point should ask a randomly chosen sensor to transmit consecutive code-letters. Finally, we consider the use of an LDPC code, which has been shown to approach channel capacity closely, and the references therein. We assume that stay-k scheduling is used. The performance of the LDPC code is simulated. It is shown that the bit error rate (BER) versus k resembles the random coding exponent versus k. Thus it makes practical sense to use the random coding exponent, which can be calculated easily, to find a good k for practical LDPC codes.

References

[1] A. Anandkumar and L. Tong, "Type-Based Random Access for Distributed Detection over Multiaccess Fading Channels," *IEEE Tran. Signal Proc.*, vol. 55, pp. 5032–5043, Oct. 2007.

[2] A. Anandkumar, J. Yukich, L. Tong, and A. Swami, "Energy Scaling Laws for Distributed Inference in Random Networks," *IEEE J. Selec. Area Comm.*, 2009. To appear and available on Arxiv `http://arxiv.org/abs/0809.0686`.

[3] A. Anandkumar, J. Yukich, L. Tong, and A. Swami, "Scaling Laws for Statistical Inference in Random Networks," in *Proc. of Allerton Conf. on Communication, Control and Computing*, (Monticello, USA), Sept. 2008.

[4] A. Anandkumar, L. Tong, and A. Swami, "Optimal Node Density for Detection in Energy Constrained Random Networks," *IEEE Tran. Signal Proc.*, vol. 56, pp. 5232–5245, Oct. 2008.

[5] Y. Sung, X. Zhang, L. Tong, and H. Poor, "Sensor Configuration and Activation for Field Detection in Large Sensor Arrays," *IEEE Tran. on Signal Processing*, vol. 56, no. 2, pp. 447–463, 2008.

[6] A. Anandkumar, L. Tong, and A. Swami, "Detection of Gauss-Markov Random Fields with Nearest-neighbor Dependency," *IEEE Tran. Information Theory*, vol. 55, pp. 816–827, Feb. 2009.

[7] A. Anandkumar, L. Tong, and A. Swami, "Distributed Estimation Via Random Access," *IEEE Tran. Information Theory*, vol. 54, pp. 3175–3181, July 2008.

[8] M. Dong, L. Tong, and B. Sadler, "Impact of MAC design on signal field reconstruction in dense sensor networks," 2007.

[9] O. Kosut and L. Tong, "The nose of a bloodhound: target chasing aided by a static sensor network," in *Proceedings of the 2007 Asilomar Conf.*, Nov 2007.

[10] O. Kosut, A. Turovsky, J. Sun, M. Ezovski, G. Whipps, and L. Tong, "Integrated mobile and static sensing for target tracking," in *Proceedings of the 2007 IEEE MILCOM*, 2007.

[11] T. He, L. Tong, and A. Swami, "On the maximum throughput of clandestine sensor networks," in *Proc. 2009 IEEE MILCOM*, (Boston), Oct 2009.

[12] Z. Yang and L. Tong, "On the Error Exponent and the Use of LDPC Codes for Cooperative Sensor Networks With Misinformed Nodes," *IEEE Transactions on Information Theory*, vol. 53, pp. 3265–3274, Sept. 2007.

[13] O. Kosut and L. Tong, "The byzantine ceo problem," in *Proceedings of the 2008 IEEE International Conference on Symposium on Information Theory*, pp. 46–50, July 2008.

[14] O. Kosut and L. Tong, “The quadratic gaussian ceo problem with byzantine agents,” in *Proceedings of the 2009 IEEE International Conference on Symposium on Information Theory*, pp. 1145–1149, 2009.